DOCUMENT RESUME

ED 460 467 EC 306 750

AUTHOR Grossen, Bonnie
TITLE How Should We Group To Achieve Excellence with Equity?
INSTITUTION National Center To Improve the Tools of Educators, Eugene, OR.
SPONS AGENCY Special Education Programs (ED/OSERS), Washington, DC.
PUB DATE 1996-07-00
NOTE 18p.
AVAILABLE FROM For full text: http://darkwing.uoregon.edu/~adiep/grp.htm.
PUB TYPE Opinion Papers (120)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Ability Grouping; Ability Identification; Court Litigation; Educational Principles; Educational Research; Elementary Secondary Education; *Equal Education; *Excellence in Education; *Grouping (Instructional Purposes); Heterogeneous Grouping; High Achievement; Homogeneous Grouping; *Instructional Effectiveness; Low Achievement; *Mixed Age Grouping; Teaching Methods; Track System (Education)

ABSTRACT

This paper addresses issues in the relationship of instructional grouping to both educational excellence and equity. First, equity issues are considered in light of two court cases: Hobson v. Hansen and Marshall v. Georgia which, together, offer guidelines for equitable ability grouping. Next, the research on the effects of grouping arrangements on learning is summarized in a table comparing four categories of grouping arrangements for dimensions, placement criteria, flexibility, instructional practices, and determination of expectations. A review on achievement grouping and tracking concludes that grouping arrangements alone are not the primary variable for school effectiveness. The importance of using effective practices for all levels, particularly the low achievement levels, is stressed. A section reviewing the research on mixed-age grouping suggests the effectiveness of achievement grouping (not mixed-age mixed-ability grouping) for instruction in reading and/or mathematics. Discussion of excellence issues focuses on use of mixed-ability grouping to achieve world class achievement. Issues addressed include similarities and differences in high quality instruction for high- and low-achieving students, use of nonstandardized expectations in high academic achievement, and mixed-ability grouping and self-esteem. A final section reviews research showing the effectiveness of "considerate" instruction versus traditional "inconsiderate" instruction for both low and high achievers. (Contains 58 references.) (DB)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 460 467

How should we group to achieve excellence with equity?

Bonnie Grossen, Ph.D

University of Oregon

July, 1996

B. Grossen

Equity Issues

The Effects of Grouping Arrangements on Learning

Research on Achievement Grouping and Tracking

Research on Mixed-Age Grouping

Excellence Issues

Can Mixed-Ability Grouping Lead to World Class Achievement?

How do we know when equity has been served?

References

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Ability grouping in America has become a loaded word. In response to inequities of the past associated with ability grouping, an emerging national agenda among nearly all reform constituencies is claiming that ability grouping is bad, it is racist, it must be eliminated (Oakes, 1985, 1990; Wheelock, 1992). Slavin (1991), for example, argues:

> "The burden of proof for the antidemocratic, antiegalitarian practice of ability grouping must be on those who would group, and no one who reads this literature could responsibly conclude that this requirement has been met." (p. 70).

Hastings sees the equity issue in more absolute terms:

> "The answer to the debate on ability grouping is not to be found in new research. There exists a body of philosophic absolutes that should include this statement: The ability grouping of students for educational opportunities in a democratic society is ethically unacceptable" (Hastings, 1992, p. 14).

Consequently, some reformers advocate not only abolishing ability grouping, but maximizing heterogeneity by mixing abilities across ages. The popular nongraded primary model of the National Association for the Education of Young Children places children most "unlike" in skill level together for instruction (Brederkamp, 1987). Reformers praise "blended" classrooms for maximizing the differences among mixed-age children in instructional groupings.

Some leaders in the international business community have a very different perspective. *The Economist* concluded in their "Education Survey" (1992) that an investor with an "eye to human capital" should look past the Anglo-Saxon world to somewhere "between the Pacific Rim and Germanic Europe." After comparing education systems around the world, the survey concludes that Germanic Europe comes out ahead because of its "unrivaled ability to churn out skilled workers." *The Economist* praises Germany's "cheerful division of schools into three kinds: grammar schools, technical schools, and vocational schools"-a tracking system designed so that, "the transition between school and work, so traumatic elsewhere, is rendered almost painless. Above all the system reinforces a culture in which training is cherished and workers revered." *The Wall Street Journal* and *Forbes* magazine have been similarly critical of currently popular American educational reforms.

In short education reformers seem to seek equity, while business seeks excellence. However, our national goal is to achieve world class excellence with equity. Equity without excellence is just as unacceptable as excellence without equity.

Equity Issues

Two Important Court Rulings

EC 306 750

BEST COPY AVAILABLE

In two landmark cases, the courts found that ability grouping resulted in a disproportionate number of minority children being placed in lower track courses. In both cases the school districts using ability grouping had the burden to prove that the grouping practices did not contribute to the differences in performance found between legally protected minority groups and white children. In other words, because disproportionately more minority children were assigned to lower groups, the defending districts had to prove that the children in the lower groups were receiving instruction that was superior to what they would otherwise achieve without ability grouping.

The decisions in these two cases were different. In *Hobson v. Hansen* (1967, 1969), the courts ruled against ability grouping. In *Marshall v. Georgia* (1984, 1985), the courts ruled in favor of ability grouping. Four critical differences made the ability grouping practices in *Marshall* (called "achievement grouping") equitable, while the ability grouping practices in *Hobson* (called "tracking") were found discriminatory and unacceptable.

1. In *Hobson*, grouping decisions were based on a measure of general ability. In *Marshall*, the level of achievement within the specific basal series was emphasized as having the most important influence on grouping decisions. "A combination of academic indicators was taken into consideration with primary emphasis being placed on a child's actual performance in the basal instructional series" (p. 18-19, Trial Opinion).

2. In *Hobson*, students were assigned to the same track for all academic instruction and these assignments remained permanent. In *Marshall*, some schools grouped children by subject and a student's assignment to high, medium, or low groups could vary depending on the subject. Furthermore, the school district provided evidence that 37% of the students in the district changed levels over the course of two academic years. Thus a student's assignment could be changed mid-stream depending on level of performance.

3. In *Hobson*, the courts found that the grouping system was associated with unequal resources and no compensatory educational benefits. In *Marshall*, the defendants claimed that because grouping decisions were based on skill levels in the basal series, greater individualization of instruction was achieved, especially at the lower levels where Chapter 1 services and the Georgia Compensatory Education Program were available.

4. In *Hobson*, no evidence was brought to show that ability grouping was having a positive effect on the learning of children in the lower tracks. In *Marshall*, the defendant school district brought evidence indicating improved performance on the Georgia Criterion Referenced Test, especially apparent for lower performing black and white students.

The *Marshall* court held that not only was ability grouping acceptable, it was preferable to mixed-ability groups because ability grouping in this case was "... designed to remedy the past results of past segregation through better educational opportunity for the present generation of black students" (p. 100). The plaintiffs offered an alternative grouping plan which called for randomly assigning students to classes. This plan was explicitly rejected by the courts as not "equally sound" (p. 26-27, Appeals Court Opinion).

In these two landmark cases, the courts distinguished between "inequitable" and "equitable" ability grouping practices. The "inequitable" ability grouping practices, called ***tracking***, involved the use of one generic score to make a permanent, comprehensive decision regarding placement with no compensatory provisions made for the lower track. The "equitable" ability grouping practices used flexible achievement groups and provided more resources for teaching children in the lower groups, which resulted in better learning. The "legal test" for equity concerned how well protected minority groups learned, not so much how they were grouped.

The Effects of Grouping Arrangements on Learning

In the recent revival of the "detracking" movement, the word "tracking" is often used to describe any form of ability grouping (Oakes, 1990; O'Neil, 1992). This broad use of the word "tracking" is misleading. Research reviewing the effects of ability grouping on learning should make the same distinctions that the courts have made between tracking and achievement grouping. Table 1 displays the critical features of important grouping arrangements.

Table 1. Features of four categories of grouping arrangements.

	Traditional	Ability Grouping		Mixed-age
Dimensions	Grouping by Age	Tracking	achievement grouping	Mixed-ability Grouping
Placement criteria	Age	IQ score or standardized general achievement score	Academic performance level in the specific subject	Mix ages, abilities, and performance levels
Flexibility	Relatively inflexible	Relatively inflexible	Changes in placement may occur at any time based on performance.	Changes in placement may occur at any time, but achievement grouping is avoided.
Instructional Practices	Vary by grade level.	Vary by track.	Matched to the level of the instructional group.	Wholistic (non-leveled), project-based, cooperative learning groups.
Expectations determined	by age level	by IQ or general achievement level	by achievement level in the specific subject	for each child by the child's teacher

Research on Achievement Grouping and Tracking

Unfortunately, the research base on grouping is extremely dated and does not clearly evaluate the four alternative grouping arrangements described in Table 1. An analysis of the dates of the most recent comprehensive reviews with opposing conclusions (Kulik and Kulik 1987; Slavin, 1987, 1990) illustrates just how dated the research is. Not one U.S. study included in Kulik and Kulik's (1987) review of 105 studies nor in Slavin's (1987, 1990) reviews of 43 elementary studies and 29 secondary studies was published after the landmark *Marshall v Georgia* ruling in 1985. Furthermore, only 5 of the 105 studies reviewed by Kulik and Kulik, and only 4 of the 72 studies reviewed in both of Slavin's reviews were published after 1976, the original passage of the Education for All Handicapped Children Act. This legislation was probably more influential than any other event in the history of American education in terms of raising the interest of school personnel in better serving the needs of students with disabilities and of low-performing students.

In fact, only 15% of the studies reviewed were published after the *Hobson v Hansen* ruling in 1969. The preponderance of the research is over 30 years old. The abuses of grouping practices that the courts called "tracking" in *Hobson v Hansen* were probably much more common across America before the Hobson ruling than they are today.

The current question of interest to schools generally differs from the researchers' questions. The researchers have generally attempted to isolate the grouping variable from instruction, keeping instruction the same for all groups and changing only the grouping arrangements. The research question is generally: Does achievement grouping improve learning when all groups are taught using the same materials and methods? Few practitioners exist who would expect achievement grouping to have any consistent effect without matching instruction to needs. The questions of interest to schools include the following:

- Is achievement grouping with appropriately varied instruction for each group more effective than mixed-age, mixed-ability grouping?
- Is achievement grouping with appropriately varied instruction more effective than traditional age-based grouping?

Research that does not attempt to vary instruction appropriately for different grouping arrangements does not answer practitioners' questions about grouping. (See Allan, 1991 and Kulik, 1991 for further details regarding the mismatch between practitioners' and researchers' questions on grouping.) Most of the studies on grouping do not describe at all the nature of the instruction that occurred in the study.

The studies of elementary school grouping alternatives have more complete descriptions of the instruction than the secondary studies of grouping. After using a "best evidence synthesis" to seek out patterns of positive and negative effects in 43 studies comparing elementary school grouping arrangements, Slavin was able to conclude:

> "Taken together, the evidence points to a conclusion that for ability grouping to be effective at the elementary level, it must create true homogeneity on the specific skill being taught and instruction must be closely tailored to students' level of performance. (p. 323)

This is consistent with the *Marshall v Georgia* ruling. The courts saw positive effects for ability grouping when the grouping was based on achievement in the specific skills taught in the program.

Furthermore, Slavin found that the conditions leading to favorable effects for grouping were more common in "within-class" grouping and rarely existed in "between-class" grouping. Within-class grouping involves assigning children to groups within a class. Between-class grouping involves assigning children to classes for the entire year based on their ability or achievement levels. Slavin reasoned that a student's placement, though optimal for one subject, may not be optimal for another in between-class grouping at the elementary level.

One model, the Joplin Plan, could not be categorized as within-class or between-class grouping. In the Joplin plan, students are grouped into mixed-age mixed-ability classes, then placed in subject-specific achievement groups formed across classes for instruction in reading and/or mathematics. For example, at a common mathematics period, all students might move to a class composed of students at the same performance level in mathematics drawn from different classes and grade levels. One mathematics group might have high first, average second, and low third graders in it, but all would be at the same approximate point in the learning sequence. These instructional groups are also flexible and not permanent. Groupings are frequently reassessed and changed if student performance warrants it. Slavin found a strong positive effect for the Joplin plan.

Based on these findings, Slavin (1991) concludes that for the elementary level he is not opposed to assigning students to mixed-ability classes and grouping children within or across classes into achievement groups when appropriate. He opposes between-class grouping where students are assigned to self-contained classes based on their ability or performance level. At the elementary level, between-class grouping approximates tracking, when the same groups are maintained for instruction in all subjects.

Slavin's (1990) review of secondary school research was more problematic. He tried again to separate the studies of within-class grouping from those of between-class grouping to determine if the same pattern of results found at the elementary level was also evident at the secondary level. He found no effects for grouping of any kind. It is not surprising that there were no effects for within-class grouping at the secondary level, though there were at the elementary level. Even if secondary teachers divided their classes into smaller groups for instruction, thereby fitting the criteria for the "within-class" grouping arrangement, it is unlikely that they would modify the instruction for each of the small groups, doubling or tripling the number of preps they would have in a day. Each group would receive only 1/3 of the instructional time they would otherwise receive.

That Slavin also found no effect for between-class (assigning students to different classes according to their achievement level) grouping at the secondary level is more surprising. Between-class grouping at the secondary level is as subject-specific as within-class grouping at the elementary level. Classes are organized by subject at the secondary level, so between-class grouping does not result in students being assigned to the same class for all subjects as it does at the elementary level.

Slavin concludes: "If the effects of ability grouping on student achievement are zero, then there is little reason to maintain the practice... Arguments in favor of ability grouping depend on assumptions about the effectiveness of grouping, at least for high achievers. In the absence of any evidence of effectiveness, these arguments cannot be sustained" (p. 492, 1990).

Slavin's (1991) suggestion that using cooperative learning with mixed-age mixed-ability groups is more viable than between-class grouping is having profound impact in the restructuring movement. (See *Educational Leadership's* issue featuring restructuring, March, 1991.) Slavin's research is frequently cited to support the extensive restructuring of secondary schools to incorporate project-based learning where small mixed-ability cooperative learning groups spend much of their school time working cooperatively on large-scale projects, such as setting up a museum featuring the local community.

However, Slavin's conclusions regarding between-class achievement grouping at the secondary level are seriously limited by the selection rules he used in his meta-analysis. Slavin systematically eliminated any study that involved different programs for different levels. Slavin included only experimental studies that compared students at the same grade level taking the same course in achievement-grouped versus nonachievement-grouped classes. For example, only ninth-grade students in Math 9 were compared. Ninth graders taking Algebra or Math 8 would not be compared with ninth-grade students taking Math 9. One treatment would involve high, average, and low sections of Math 9. The other treatment involved all levels mixed in Math 9 classes. Slavin comments regarding this limitation:

> "The experimental studies do not compare students in Algebra 1 to those in Math 9, or students who take 4 years of math to those who take 2. The conclusions drawn in this section are limited, therefore, to the effects of between-class grouping *within the same courses*, and should not be read as indicating a lack of differential effects of tracking [or achievement grouping]. (Slavin, 1990, p. 486-7)

This is a major caveat. Most of the practical impact of achievement grouping would be expected to come from high level students taking courses that cover more advanced content. Any studies that would detect this effect were excluded from Slavin's reviews.

Kulik and Kulik (1991) used different selection criteria for their metaanalyses and ended up including a different set of studies. Very few studies reviewed by Slavin were also reviewed by the Kuliks. In discussing the results of the Kulik and Kulik review (1991), Kulik (1991) distinguished three types of programs:

> Type I: simple programs in which all ability groups are taught with the same or similar materials and by the same or similar methods.
> Type II: programs in which teaching materials and methods are adjusted to meet the special needs of a specific aptitude group (for example, enriched instruction for the talented and gifted).
> Type III: programs in which adjustment of teaching materials is so extensive that it affects a student's rate of progress through school (for example, programs of accelerated instruction).

Effects varied according to type, with negligible effects found for Type I programs (.1 effect size), stronger effects for Type II programs (.4 effect size), and much stronger effects for Type III programs (1.0 effect size). Kulik's (1991) conclusions seem to support the practice of achievement grouping as defined by the courts. The more instruction is varied to meet the specific needs of students in the achievement groups, the more effective it is.

However, most of the Type II and Type III research evaluated only programs for the gifted and high-performing students. As Slavin (1991) points out, evaluating the effects of gifted programs only on gifted students leaves open the possibility that gifted programs might have positive effects for all students. Indeed many reformers (e.g., Oakes; see interview with Oakes in O'Neil, 1992) argue that gifted programs should be offered to all students. However, the effectiveness of gifted programs for all students was not evaluated in this research. Other research (described later) raises considerable doubt that gifted programs would have positive effects for all students.

Summary. Flawed research methodology seems to support the conclusion that there is no clear answer to the question: Does achievement grouping improve learning when all groups are taught using the same materials and methods? This is a question few ask. The contradictions in the findings within each

metaanalysis seem to indicate that grouping arrangements alone are not the primary variable for school effectiveness. Whether effective practices are used for all levels, particularly the low achievement levels, is the legal test for racial equity. If the learning of low-achieving minority children is accelerated, equity is served. If not, inequity is present.

Research on Mixed-Age Grouping

Pavan (1977) reviewed 51 comparisons of mixed-age grouping conducted between 1968 and 1978 and concluded that *mixed-age* grouping was more effective than *age-based* grouping. Pavan's conclusion was used to support the nongraded model promoted by the National Association for the Education of Young Children (Brederkamp, 1987), which not only mixes ages, but also mixes abilities. However, Pavan's research does not support mixed-ability grouping within the mixed-age model. The mixed-age models she evaluated included both achievement grouping, as in the Joplin plan, and mixed-ability grouping. Pavan did not break down the results for mixed-age models according to whether achievement grouping or mixed-ability grouping was used. Rather she grouped the effects together.

Gutiérrez and Slavin (1992) reviewed Pavan's same data set and more (57 studies), but categorized the studies according to instructional and grouping practices used among the mixed-age models. Their findings did not contradict Pavan's; they also found more positive than negative significant results favoring the mixed-age ("nongraded") model. However, they found that the models that contributed most to the overall positive effect Pavan found for mixed-age primaries actually used achievement grouping for instruction in reading and/or mathematics (the Joplin plan), not mixed-age mixed-ability grouping, as is promoted by Pavan (1992) and Brederkamp (1987). Gutiérrez and Slavin concluded that the "nongraded organization can have a positive effect on student achievement if cross-age grouping is used to allow teachers to provide more direct instruction to students but not if it is used as a framework for individualized instruction" (p. 333).

Achievement grouping across ages, rather than only within grade levels, allows teachers to reduce the number of within-class reading and math groups they teach at any given time, thereby reducing the need for independent seatwork and follow-up. Gutiérrez and Slavin (1992) indicated that several evaluators of Joplin-like programs noted specifically that mixed-age groupings made within-class groupings unnecessary, so teachers could use the entire class period to teach the whole class. Mixed-ability models involved individualized instruction, learning stations, learning activity packets, and other individualized or small group activities which reduced direct instruction time with little corresponding increase in appropriateness of instruction to meet individual needs, according to Gutiérrez and Slavin (1991). They point out that the research on nongradedness has not evaluated the currently popular model promoted by the NAEYC and Katz et al. (1991):

> The movement toward developmentally appropriate early childhood education and its association with nongrading means that the nongraded primary schools of the 1990s will often incorporate 4- and 5-year-olds (earlier forms rarely did so) and that instruction in nongraded primary programs will probably be more integrated and thematic, and less academically structured or hierarchical, than other schools.... Whether these models will have positive or negative effects on ultimate achievement is currently unknown. (p. 370)

Anderson and Pavan (1993) later expanded Pavan's original review (1977) of nongraded, or mixed-age primaries, to include 64 studies. They found positive effects for the nongraded model, but again they did not break down the results according to whether the models used mixed-ability or achievement grouping within the mixed-age model. Without this breakdown, their conclusions cannot be used to support mixed-ability grouping practices within the mixed-age model.

Gutiérrez and Slavin (1992) also point out an additional problem with the research on nongraded models: If the nongraded model is used to allow students more time to complete the primary grades, as they usually are, then the average "third-year" student may be older in the nongraded school than in the graded school, creating an artificial advantage for the nongraded model in this research literature.

McGurk and Pimentle (1992) also found empirical support for the Joplin plan in their review of the

research on mixed-age (nongraded) models. Mixed-age models that did not use the Joplin plan obtained academic achievement that was comparable to the age-based grouping. Pratt (1986) found no consistent advantage for one grouping plan over another in academic achievement, nor did Cotton (1993), Miller (1990; 1991), and Ford (1977). In their review of reviews, Ellis and Fouts (1994) conclude that most reviews find the nongraded primary has no positive effects on achievement.

Summary. The research on mixed-age models includes mixed-ability and achievement grouping within a mixed-age environment. The findings cannot be used to understand the effects of achievement or mixed-ability grouping without separate analysis. Separate analyses indicate that better results are associated with the Joplin plan for achievement grouping. An important question left unanswered in all of these reviews is how well the low-performing students did. As the courts have already ruled, the question is not whether a school groups by ability or not; the question is how well the low-performers do, especially when they include a larger proportion of legally protected minority students. If these low achieving students are not learning as well as they could, equity is not being served, regardless of the grouping arrangement.

Excellence Issues

Our national reform goal is to achieve world class standards. A key recommendation of many organizations leading our national reform efforts is to achieve equity by mixing students with widely differing abilities in the classroom. Achieving world class standards though requires much more. Another approach to resolving the problem of equity is to look for school models where low achievers reach remarkably high performance levels and find reliable ways to replicate those models.

One of the few organizations that has taken a serious look at identifying the best performance in the world is the American Federation of Teachers (AFT). A recent comparison of the achievement levels of lower track students in European countries with American students reveals that lower track students in Europe achieve remarkably high performance levels compared to mainstream students in America (AFT, 1995). The gateway exams for school completion for lower track students in Europe are much more rigorous than America's comparable exam for a Graduation Equivalency Diploma, which is normed to reflect what 75% of America's high school graduates know by the end of grade 12. At grade 9 or grade 10, 60% to 85.5% of the students in European countries pass their much more rigorous exams. The achievement levels of lower track students in European countries using tracking systems are much higher than the expectations for American students.

Certainly the relatively homogeneous societies of Europe do not face the same equity issues that the racially heterogeneous American society faces. If transferred to America, the more rigid tracking of students into different schools at an early age and the permanent assignment of students to classroom groups over several years could easily translate into permanently lower expectations for minority children.

Tracking per se is not necessarily the cause of the high performance levels for lower track students in Europe. The American Federation of Teachers suggests other factors leading to the effectiveness of the European system: national or state-administered assessments, strong incentives to excel, and a common curriculum. These aspects of the European model seem crucial if world class excellence is to be achieved.

Can Mixed-Ability Grouping Lead to World Class Achievement?

If mixed-age mixed-ability grouping can result in low achievers reaching the same high performance levels found in Europe, then achievement grouping is not necessary. The fact that this challenge has not been met using mixed-age mixed-ability grouping does not mean that the challenge is impossible to meet. However, there are several requirements that mixed-age mixed-ability grouping must meet in order to make the case that world class excellence can be achieved using mixed-ability grouping.

Does quality instruction look the same for high- and low-achieving students? Mixed-ability grouping assumes that the same kind of instruction is best for achieving excellence with both high and low

achievers. In her frequently cited book, *Keeping Track*, Oakes (1985) analyzed descriptive data collected on 25 secondary schools during the early 1970's and documented that inferior instruction was still occurring in many schools, in spite of the 1967 and 1969 *Hobson v Hansen* rulings. She judged the instruction for the low groups inferior not because fewer resources were available to these groups, as the courts did. She judged the instruction in the low groups inferior because the quality of instruction was different. Low groups did lots of worksheets, worked alone more, and spent more time reading out of textbooks. The high groups received more experience-based learning and challenging problems that are likely to have more than one right answer (O'Neil, 1992).

Oakes argues that with mixed-ability grouping, all students will have equal access to the higher quality instruction. Her argument assumes that what she has identified as "quality" instruction will have the same beneficial results for both high and low-achievers. Only under this condition is equity achieved by providing the same instruction for all students.

A very recent study by Gamoran, Nystrand, Berends, and LePore (1995) evaluated the effects of various instructional variables on the learning of high and low performing students. They examined the characteristics of students placed in 92 honors, regular, and remedial English classes in eighth and ninth grade, looking at the effects of similarities and differences in the instruction across achievement groups on the learning of these groups. They found that some instructional variables¾discussion and authentic questions¾had reversed effects on the achievement of different achievement groups:

> "This difference [in the levels of discussion across groups] turned out to be potent for achievement inequality, however, because discussion only benefited students in the high-level classes. Authenticity was also consequential for achievement gaps, but not in the way originally expected: It occurred with similar frequency across classes, but it was beneficial to high-ability students and detrimental to those in low-ability classes." (p. 708)

The finding for discussion "contradicted our expectation that discussion would benefit low-ability students most of all" (p. 706). The finding for authenticity was "not consistent ... with our speculation, based on prior research, that authentic discourse offers greater benefits in low-ability classes than elsewhere. We found just the opposite" (p. 706).

Gamoran et al.'s study (1995) is important because it raises a crucial question: Does quality instruction look the same for high- and low-ability students? If features of quality vary according to the achievement level of the group, then Oakes (1985), and similarly Goodlad's (1984), argument is flawed. What these researchers thought was a feature of high-quality instruction (authentic questions, open-ended discussion) may actually not represent high quality instruction for students at lower achievement levels. Mixing low achievers with high achievers and providing instruction that benefits only high achievers could have the opposite effect and not increase equity.

Can nonstandardized expectations result in world class achievement? Expectations play an important role in achievement (Means, Moore, Gagne, & Hauck, 1979; Rist, 1970). Different grouping arrangements have strong implications for student expectations. In three of the four models in Table 1, age-based grouping, tracking, and achievement grouping, expectations can be clearly defined, or standardized, for each group. In mixed-age, mixed-ability grouping, common expectations do not exist for the group, but vary by individual.

When students are grouped by age, all children of the same age face the same grade-level standards and are expected to learn the curriculum provided for that grade level. Early proponents of tracking criticized the appropriateness of age-based expectations (Turney, 1931), just as current advocates of mixed-age, mixed-ability grouping do (Brederkamp, 1987). Not all children of the same age should be expected to achieve the same outcomes. Tracking redefines expectations for a child's performance based on the child's general ability rather than age. Expectations though are still standardized for the different tracks (e.g., European systems).

Achievement grouping temporarily redefines short term expectations based on the current achievement level of the child in the specific subject. All children in a given achievement group generally start from

the same place, with different achievement groups starting from different places. Long-term expectations though are generally referenced to the age-level expectations. All achievement groups within the same larger class group work toward achieving, at a minimum, the same long-term expectations defined for that group. Some achievement groups may exceed these standardized expectations.

In mixed-age, mixed-ability grouping expectations vary by individual. The teacher is the judge of what should be expected of each individual and the children are not pressured to achieve expectations that are inappropriate for them (Brederkamp, 1987). In theory varied expectations for each individual sounds fair and equitable. In reality though, does it work out that way? How does mixed-ability grouping with variable expectations interact with the noted tendency that teachers tend to communicate more positively with children they perceive as bright and more negatively with children they perceive as slow (Cooper, 1979).

Some ethnographic research evaluated the fairness of teachers in varying expectations appropriately in "progressive" schools that emphasized the importance of variable expectations according to the unique abilities of each child (Atkinson, 1985; Bernstein, 1974; Sharp, Green, & Lewis, 1975; Simon, 1981; Willis, 1977). Atkinson (1985) concluded that the shift from traditional to progressive methods in England represented a shift from visible to invisible control.

Sharp, Green, and Lewis (1975) describe how this shift occurs in case studies of three teachers in a model progressive school:

> "Whereas all three teachers would claim to be supporters of the egalitarian principle that all pupils are of equal worth, having an equal right to receive an education appropriate to their needs, in practice there was a marked degree of differentiation among the pupils in terms of the amounts and kinds of interaction they had with their teachers....Those pupils whom their teachers regarded as more successful tended to be given far greater attention than the others. The teachers interacted with them more frequently, payed [sic] closer attention to their activities, subtly structuring and directing their efforts in ways which were noticeably different from the relationship with other pupils less favourably categorized." (p. 115)

The children who received less attention were the lower performing children who were from lower working class families, while the children the teacher spent more time with were higher performing children who were also from a higher social class. These inequities occurred in classrooms using mixed-ability grouping taught by teachers espousing strong beliefs in the egalitarian principles undergirding progressivism.

For example, Michael's teacher described him as a "peculiar" boy who wants to "go his own sweet way." The teacher said she would not "force" or "make" Michael do activities, even where his achievement was poor compared with other children, because to do so would violate the integrity of the child. Yet she did say: "But he's ever so willing to join in if you organize a little group-but he doesn't *need* to...," so Michael often was not invited to participate (pp. 137-8, Sharp, Green, & Lewis, 1975).

Similar observations were made by other ethnographic researchers, who also shared the egalitarian goals of progressivism (Atkinson, 1985; Bernstein, 1974; Simon, 1981; Willis, 1977). For example, Willis (1977) concluded:

> "...it can be argued that often "progressivism" has had the contradictory and unintended effect of helping to strengthen processes within the counter-school culture which are responsible for the particular subjective preparation of labour power and acceptance of a working class future in a way which is the very opposite of progressive intentions in education." (p. 178)

Apparently, holding different expectations for different students in the same instructional groups, as is recommended in mixed-age mixed-ability grouping arrangements, can result in a much more insidious form of inequality. When the same expectations are held for all members of the group, as occurs in achievement grouping or age-based grouping arrangements, and even in tracking, the differential

expectations for the different groups are at least public and can be agreed upon in a partnership of teachers, parents, and children. The openness of the expectations for each group is possibly more democratic than the veiled nature of a teacher's arbitrary, personal expectations for each student in a mixed-age mixed-ability group. At least, one certainly cannot simply assume that equity will be better served by mixed-age, mixed-ability grouping.

An important point that seems often overlooked is that a model that emphasizes variable expectations for each individual student is also incompatible with our national goal to establish standards. In reconciling the NAEYC's nongraded, mixed-ability model, which emphasizes developmentally appropriate expectations, with the national movement to establish standards, the NAEYC advocates that governing bodies redefine standards to mean not what students should be able to do, but how teachers should teach.

Does mixed-ability grouping raise self-esteem? If it does, the next question is whether higher self-esteem significantly contributes to excellence. A major criticism of achievement grouping is that it lowers the self-esteem of students in low-achievement groups. Kulik and Kulik (1982) and Kulik (1985) reviewed the research regarding effects of grouping on attitude and self-esteem. They found that achievement grouping in a subject resulted in a better attitude toward that subject but did not change attitudes about school.

In regard to self-esteem, the Kuliks' findings contradict the prevailing expectation. Achievement grouping into high, average, and low groups had a small overall effect on self-esteem, but effects tended to be slightly positive for low-achievement groups and slightly negative for high and average ones (Kulik & Kulik, 1982; Kulik, 1985). Limited studies of remedial programs indicate that achievement grouping has positive effects on the self-esteem of slow learners (Kulik, 1985). Vaughn (in press) has found similar results in a longitudinal study. Self-esteem decreased for children who moved from the low achievement group into mixed-ability classes.

Allan (1991) asked Kulik for a possible explanation for this surprising result:

> "Kulik (personal communication) raises an interesting point on the relative importance of the effects of labeling versus the effects of daily classroom experience. He suggests that the labeling (by placement of a student into a low-medium-high group) may have some transitory impact on self-esteem but that impact may be quickly overshadowed by the effect of the comparison that the student makes between himself or herself and others each day in the classroom. Low-ability students may experience feelings of success and competency when in a classroom with others of like ability, and high-ability students may encounter greater competition for the first time. While the data cannot, in themselves, identify the cause of these findings, the results make it clear that we must reexamine the arguments about self-esteem in light of them." (p. 64)

Other research is often cited to contradict these conclusions. Analyses of the effects of the nongraded primary on self-esteem and attitude frequently find that the nongraded primary has positive effects on both (Ford, 1977; Johnson, Johnson, Pierson, & Lyons, 1985; Miller, 1990; Pavan, 1977; Pratt, 1986; Way, 1981). However, as noted earlier, the nongraded model has included both mixed-age achievement grouping, as in the Joplin plan, and mixed-age mixed-ability grouping. The findings do not necessarily indicate that the models that mixed abilities caused these effects.

In the evaluation of Project Follow Through, the largest educational study ever funded by the U.S. Department of Education, Abt Associates reported very surprising results for self-esteem (1977). The most effective model, which used achievement grouping, produced the largest effects for self-esteem, indicating that self-esteem may be more a function of successful learning than grouping arrangement.

> "The performance of Follow Through children in the Direct Instruction sites on the affective measures is an unexpected result. The Direct Instruction Model does not explicitly emphasize affective outcomes of instruction, but the sponsor has asserted that they will be consequences of effective teaching. Critics of the model have predicted that the emphasis on tightly controlled instruction might discourage children from freely expressing themselves,

and thus inhibit the development of self-esteem and other affective skills. In fact, this is not the case." (Abt, IV-B, 1977, p. 73)

The five major models evaluated in Project Follow Through claiming self-esteem as an important goal actually resulted in more negative effects for self-esteem when compared to traditional models of schooling.

How do we know when equity has been served?

To argue that separating children by achievement levels denies them equity in education assumes that the classroom is much like a bus: If students have equal access to a seat in the classroom, equity has been served. Equity in education requires more. Equity is clearly served when the achievement of minority children matches the best achievement in the world. Equity is clearly served when the growth rates of children starting at low achievement levels matches or exceeds the growth rates of children starting at high achievement levels. By observing closely when these events occur, educators may learn more about what it takes to achieve excellence with equity. The critical variables have more to do with instruction than with grouping.

Minority children have achieved at world class levels. The Center for the Development and Study of Effective Pedagogy for African-American Learners (CPAL) at Texas Southern University has identified elementary schools in Texas that have achieved remarkable levels with economically disadvantaged African-American children. Pietsch Elementary in Beaumont, Texas, was one of few schools to receive an "Exemplary" rating for the performance of their low income African-American children on the Texas Assessment of Academic Skills in 1995. An "Exemplary" rating is given to schools in which 90% of the African-American students meet all the state standards in reading, writing, and mathematics. A rating of "Recognized" was given to schools with 70% of the students meeting the standards and rating of "Acceptable" is given when only 25% of the students meet the standards.

Most schools in Texas achieve a rating of "acceptable." At Pietsch though, 94% of African-American students the met the standards in reading; 92% in mathematics. Among the Hispanic students at Pietsch, 90% passed the standards for reading and 100% passed the standards for mathematics. Three years ago, Pietsch Elementary students were performing around the 20th percentile. The principal attributes their recent success to the implementation of the University of Oregon Direct Instruction model three years ago.

Table 2. The Contrast Between Considerate Instruction and Traditional Inconsiderate.

Considerate	Traditional Inconsiderate
Present **Big Ideas**, concepts and principles that facilitate the most efficient and broad acquisition of knowledge across a range of examples. Big ideas make it possible for students to learn the most and learn it as efficiently as possible, because "small" ideas can often be best understood in relationship to larger, "umbrella concepts."	Present a barrage of unrelated facts and details. The links between concepts are obscured.
Teach **Conspicuous Strategies**, which are made up of specific steps that lead to solving complex problems.	Strategies are seldom taught.
Mediated Scaffolding provides personal guidance, assistance, and support.	Little direction or provision for scaffolding the progression of learning toward greater independence is provided.
Strategic Integration of new knowledge with old knowledge.	Spiraling of topics does not carefully integrate concepts.
Background Knowledge is pretaught.	Important prerequisite learning is often not evaluated nor taught.
Judicious Review requires students to draw upon and apply previously taught knowledge over time.	Review is often minimal.

Kreole Elementary in Moss Point, Mississippi, had a history of scoring around the 20th percentile on state standardized tests of reading and mathematics. After implementing the University of Oregon Direct Instruction Model, Kreole Elementary made headline news March 29, 1995 for scoring second highest in fourth-grade reading in Mississippi. Students averaged the 87th percentile in reading and the 79th percentile in mathematics in 1994. The fourth-grade pupils scored tenth highest in language arts. This achievement is so remarkable because the children of Kreole Elementary are 85% "poverty-level," African-American children.

Barclay Elementary serves a largely low-income (82% free lunch), African-American population in Baltimore. Barclay students scored consistently below the 40th percentile before implementing the Calvert model. During each of the three successive years of using the Calvert model, Barclay pupils' scores were higher than the year before. Referrals to Chapter 1 and Special Education have dropped by more than half, and referrals to the district's Gifted and Talented Education program have risen dramatically (Stringfield, 1995). Stringfield's (1995) evaluation concludes that "the striking results derive from the adoption of a very well designed, highly demanding, continuously evaluated curriculum and instructional program, and a set of highly reliable implementation techniques" (p. 1). All three of these high-achieving schools use achievement grouping during at least part of the school day.

Low performing children have learned at remarkable rates and achieved at remarkable levels. Remarkable achievement levels for students with disabilities have also been obtained. The National Center to Improve the Tools of Educators (NCITE) has synthesized empirical research to identify the critical features of instruction that accelerates the achievement of diverse learners (children of poverty, children with limited English, and children with disabilities). We have called this instruction "considerate" because it improves learning by placing greater effort into the design of the instructional activities (Grossen & Carnine, in press). Table 2 contrasts considerate instruction with traditional instruction.

The features of considerate instruction align closely with the instructional models used in the high-performing schools described above. Considerate instruction seems effective with children with disabilities as well as with children of poverty for several reasons. The barriers that disabilities and poverty bring to achievement seem to limit the academically relevant background knowledge that children bring to school. Considerate instruction works to overcome this by assuming nothing without evaluating whether children have the prerequisite knowledge to succeed in a specific instructional unit. Efficiently providing children with relevant background knowledge seems crucial to their future learning.

Some of the results that have been achieved with students with disabilities in experimental studies evaluating considerate instruction are highlighted in Table 3. Many of the studies in Table 3 involved mainstreamed students with learning disabilities receiving instruction with general education students. Generally, we have found that mixing students with disabilities with general education students is most effective when the content of the instruction is new for all students. For example, the considerate earth science instruction started by assuming the children knew nothing about earth science. In most cases, general education students know as little about earth science as students with disabilities. So in this case, grouping different abilities of students together was effective, because all were starting with a relatively equal knowledge base in science.

Not all of our work with special education students working in the mainstream has been as effective. For example, our work teaching reasoning to nonmainstreamed students with learning disabilities was quite effective when these students were

grouped separately (see 1 and 2 below). However, when we used the intervention with mainstreamed students with disabilities, they achieved only very meager outcomes in the same amount of time using the same intervention and measures. The instruction seemed to benefit average and high-performing students much more (Grossen, Lee, & Johnson, 1996). In the area of reasoning, the students with disabilities did not start at the same achievement level. Facilitating the needs of students who are missing some basic reasoning skills in the same classroom with students who were not missing those skills seems to reduce the amount of appropriate instruction the lower performing students receive.

In a two-year study of mathematics, we found that mainstreamed students with disabilities did well both years. During the second year, approximately one-third of the class was new. These students, though they came from general education settings, did not have the same background in mathematics that the original group had. It was far more difficult for the teacher to meet the needs of these new general education students, than it was for her to continue meeting the needs of the students with disabilities. In fact, 3 of the 5 students with disabilities became classroom "stars" during the second year, often providing tutoring for the general education students who were new to the class.

Table 3. Research on the Effects of Considerate Instruction In Closing the Gap Between Special Education and General Education Students.

Reasoning

1. On a variety of measures of argument construction and critiquing, achievement-grouped high school students with learning disabilities scored as high as or higher than high school students in an honors English class and college students enrolled in a teacher certification program (Grossen & Carnine, 1990).

2. In constructing arguments, achievement-grouped high school students with disabilities scored significantly higher than college students enrolled in a teacher certification program and scored at the same level as general education high school students. All of these groups had scores significantly lower than those of the college students enrolled in a logic course (Collins & Carnine, 1988).

Science

3. On a test of problem solving to achieve better health, achievement-grouped high school students with disabilities scored significantly higher than nondisabled students who had completed a traditional high school health class (Woodward, Carnine, & Gersten, 1988).

4. On a test of problem solving that required applying theoretical knowledge and predicting results based on given information, mainstreamed middle school students with disabilities scored higher than a class of general education students taught in a student-centered treatment (Grossen, Carnine, & Lee, 1996).

5. On a test of misconceptions in earth science, mainstreamed middle school students with learning disabilities showed better conceptual understanding than Harvard graduates interviewed in Schnep's 1987 film, *A Private Universe* (Muthukrishna, Carnine, Grossen, & Miller, 1993).

6. On a test of earth science problem solving, mainstreamed middle school students with learning disabilities scored significantly higher than nondisabled students who received traditional science instruction (Woodward & Noell, 1992).

7. On a test of problem solving involving earth science content, most of a group of mainstreamed middle school students with learning disabilities scored higher than the mean score of the nondisabled control students (Niedelman, 1992).

Mathematics

8. On a test of problem solving requiring the use of ratios and proportions, mainstreamed high school students with disabilities scored as well as nondisabled high school students who received traditional math instruction (Moore & Carnine, 1989).

9. On a test requiring the application of fractions, decimals, and percents, age-grouped fifth and sixth grade low-achieving students scored significantly higher than high-achieving students in a constructivist treatment (Grossen & Ewing, 1996).

History

10. On a history test that required analyzing primary source documents, the scores that mainstreamed high school students with learning disabilities attained on the use of principles and facts in writing did not differ significantly from nondisabled control students (Crawford & Carnine, 1994).

Based on NCITE's research it seems that achievement level is a crucial consideration in providing highly effective instruction. General ability level is much less important, if considerate instruction is used. With considerate instruction, low achieving children are capable of achieving at remarkable levels, regardless of whether the low achievement is due to disabilities in the child or due to economic deprivation.

Conclusion

To move from achievement grouping to mixed-age grouping because low achievers have not been successful in achievement groups (e.g., Evans, 1991; Slavin, 1990) is not sufficient to achieve equity. The courts determined in *Marshall v Georgia* that to establish equity, the performance of low achieving groups must improve. If low achievers remain unsuccessful in mixed-ability classes, equity is still not achieved. The research cited in support of dismantling achievement grouping systems at best finds that the effects of achievement and mixed-ability grouping are the same (Slavin, 1990). The implication of this research is that low achievers will likely remain unsuccessful in "detracked" schools. The challenge remains for schools to improve the achievement levels of these low achieving children. There is no equity without excellence.

Several models demonstrate what traditionally low-performing groups of children are capable of achieving, both children of poverty and children with disabilities. All of these models incorporate a well designed, highly demanding, continuously evaluated curriculum and instructional program, and a set of highly reliable implementation techniques. The search for equity cannot ignore these results.

Visit NCITE's web page for more information at http://darkwing.uoregon.edu/~ncite/

References

Abt Associates. (1977). *Education as experimentation: A planned variation model* (Vol. 4B Effects of follow through models). Cambridge, MA: Author.

Allan, S (1991). Ability-grouping research reviews: What do they say about grouping and the gifted? *Educational Leadership, 48*(6), 60-65.

Anderson, R.H., & Pavan, B.N. (1993). *Nongradedness: Helping It to Happen*. Lancaster, PA: Technomic Press.

Atkinson, P. (1985). *Language, structure and reproduction: An introduction to the sociology of Basil Bernstein*. London: Methuen & Co.

Bernstein, B. (1975). *Class, codes and control* (Vol. 3: Towards a theory of educational transmissions). London: Routledge & Kegan Paul.

Brederkamp, S. (Ed.). (1987). Developmentally appropriate practice in early childhood programs serving children from birth through age 8. Washington, DC: National Association for the Education of Young Children.

Brewer, D., Rees, D., & Argys, L. (1995). Detracking America's schools: The reform without cost? *Phi Delta Kappan, 77(3)*, 210-215.

Collins, M., & Carnine, D. (1988). Evaluating the field test revision process by comparing two versions of a reasoning skills CAI program. Journal of Learning Disabilities, 21, 375-379.

Cooper, H. (1979). Pygmalion grows up: A model for teacher expectations, communication and performance influence. *Review of Educational Research, 49*(3), 389-410.

Cotton, K. (1993). *Nongraded primary education*. Portland, OR: Northwest Regional Educational Laboratory.

Crawford, D., & Carnine, D. (1996). Promoting and assessing higher order thinking in history: Using performance assessment to evaluate effects of instruction. (Technical Rep. 101). Eugene, OR: National Center to Improve the Tools of

Educators, University of Oregon.

Ellis, A., & Fouts, J. (1994). *Research on school restructuring*. Princeton, NJ: Eye on Education.

Evans, D. (1991). The realities of un-tracking a high school. *Educational Leadership, 48*(8), 16-17.

Ford, B. (1977). Multiage grouping in elementary school and children's affective development: A review of recent research. *The Elementary School Journal, 78*(2), 149-159.

Gamoran, A. (1992). Is ability grouping equitable? *Educational Leadership, 50*(2), 11-17.

Gamoran, A., & Mare. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality? *American Journal of Sociology, 94*, 1146-1183.

Gamoran, A., Nystrand, M., Berends, M., & LePore, P. (1995). An organizational analysis of the effects of ability grouping. *American Educational Research Journal, 32*(4), 687-715.

Goodlad, J. (1984). *A place called school*. New York: McGraw-Hill.

Grossen, B., & Carnine, D. (1990). Diagramming a logic strategy: Effects on more difficult problem types and transfer. Learning Disability Quarterly, 13, 168-182.

Grossen, B., & Carnine, D. (1996). Considerate instruction helps students with disabilities achieve world class standards. *Teaching Exceptional Children, 28*(4), 77-81.

Grossen, B., & Ewing, S. (1994). Raising mathematics problem-solving performance: Do the NCTM teaching standards help? (Technical Rep. 102). Eugene, OR: National Center to Improve the Tools of Educators, University of Oregon.

Grossen, B., Carnine, D., & Lee, C. (1996). The effects of considerate instruction and constructivist instruction on middle-school students' achievement and problem solving in earth science. (Technical Rep. 103). Eugene, OR: National Center to Improve the Tools of Educators, University of Oregon.

Grossen, B., Lee, C., & Johnson, D. (1996). A comparison of the effects of considerate instruction in reasoning with constructivism on deductive reasoning. (Technical report). Eugene, OR: National Center to Improve the Tools of Educators, University of Oregon.

Gutiérrez, R., & Slavin, R. (1992). Achievement effects of the nongraded elementary school: A best-evidence synthesis. *Review of Educational Research, 62*(4), 333-376.

Hastings, C. (1992). Ending ability grouping is a moral imperative. *Educational Leadership, 50*(2), 14.

Hobson v Hansen, 269 F. Supp. 401 (D.D.C.,1967). Affirmed, *Smuck v Hobson*, 408 F. 2d 175 (D.C. Cir, 1969).

Johnson, D., Johnson, R., Pierson, W., & Lyons, V. (1985). Controversy versus concurrence seeking in multi-grade and single-grade learning groups. *Journal of Research in Science Teaching, 22*(9), 835-848.

Katz, L.G., Evangelou, d., & Hartman, J.A. (1991). *The case for mixed-age grouping in early childhood education*. Washington, DC: National Association for the Education of Young Children.

Kavale, K.A. (1987). Introduction: Effectiveness of differential programming in serving handicapped students. In M.C. Wang, M.C. Reynolds, and H.J. Walberg (Eds.), *Handbook of special education: Research and practice, learner characteristics and adaptive education* (Vol. 1). Oxford, England: Pergamon Press.

Kulik, C.-L. (1985). *Effects of inter-class ability grouping on achievement and self-esteem*. Paper presented at the annual convention of the American Psychological Association (93rd), Los Angeles, California.

Kulik, J.A., & Kulik, C.C. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal, 19*, 415-428.

Kulik, J.A., & Kulik, C.C. (1987). Effects of ability grouping on students achievement. *Equity & Excellence, 23*(1-2), 22-30.

Kulik, J.A. (1991). Findings on grouping are often distorted: Response to Allan. *Educational Leadership, 48*(6), 67.

Marshall et al. v Georgia. U.S. District Court for the Southern District of Georgia, CV482-233, June 28, 1984; Affd (11th cir.no. 84-8771, October 29, 1985). Note, the Court of Appeals decision was published as *Georgia State Conference of Branches of NAACP v State of Georgia.*

McGurk, E., & Pimentle, J. (1992). *Alternative instructional grouping practices.* (ERIC Document Reproduction Service No. ED353279).

Means, V., Moore, J., Gagne, E., & Hauck, W. (1979). The interactive effects of consonant and dissonant teacher expectancy and feedback communication on student performance in a natural school setting. *AERA Journal, 16*(4), 367-374.

Miller, B. (1990). A review of the quantitative research on multigrade instruction. *Research in Rural Education, 7*(1), 1-8.

Moore, L.., & Carnine, D. (1989). Evaluating curriculum design in the context of active teaching. Remedial and Special Education, 10, 28-37.

Muthukrishna, N., Carnine, D., Grossen, G., & Miller, S. (1993). Children's Alternative Frameworks: Should They Be Directly Addressed in Science Instruction? Journal of Research in Science Teaching, 30(3), 233-248.

Niedelmann, M. (1992). Problem solving and transfer. In D. Carnine & E. Kameenui (Eds.), Higher order thinking: Designing curriculum for mainstreamed students (pp. 137-156). Austin TX: Pro Ed.

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven, CT: Yale University Press.

Oakes, J. (1990). *Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science.* Santa Monica: Rand.

O'Neil, J. (1992). On tracking and individual differences: A conversation with Jeannie Oakes, *Educational Leadership, 50*(2), 18-21.

Pratt, D. (1986). On the merits of multiage classrooms. Their work life. *Research in Rural Education, 3*(3), 111-116.

Rist, R. (1970). Student social class and teacher expectations: A self-fulfilling prophecy in ghetto education. *Harvard Educational Review, 40*(3) 411-451.

Sharp, R., Green, A., & Lewis, J. (1975). *Education and social control: A study in progressive primary education.* London: Routledge & Kegan Paul.

Simon, B. (1981). The primary school revolution: Myth or reality? In B. Simon & J. Willcocks (Eds.), *Research and practice in the primary classroom.* London: Routledge & Kegan Paul.

Slavin, R. (1991). Are cooperative learning and "untracking" harmful to the gifted? Response to Allan. *Educational Leadership, 48*(6), 68-71.

Slavin, R. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research, 57*(3), 293-336.

Slavin, R. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research, 60*(3), 471-499.

Stringfield, S. (1995). Fourth year evaluation of the Calvert school program at Barclay school. Baltimore, MD: Center for the Social Organization of Schools, Johns Hopkins University.

Turney, A. H. (1931). The status of ability grouping. *Educational Administration and Supervision, 17,* 21-42, 110-127.

Vaughn, S. (in press). University of Miami.

Way, J. (1981). Achievement and self-concept in multiage classrooms. *Educational Research Quarterly, 6*(2), 69-75.

Wheelock, A. (1992). *Crossing the tracks: How "untracking" can save America's schools*. New York, NY: The New Press.

Willis, P. (1977). *Learning to labour: How working class kids get working class jobs*. Westmead, London: Saxon House.

Woodward, J., & Noell, J. (1992). Science instruction at the secondary level: Implications for students with learning disabilities. In D. Carnine & E. Kameenui (Eds.), *Higher order thinking: Designing curriculum for mainstreamed students* (pp. 39-58). Austin TX: Pro Ed.

Woodward, J., Carnine, D., & Gersten, R. (1988). Teaching problem solving through a computer simulation. *American Educational Research Journal, 25*(1), 72-86.